

Una encuesta como un recurso docente integrador para una asignatura de estadística en estudios técnicos universitarios

Marta Ginovart¹, Mónica Blanco¹, y Xavier Portell²

¹Dept. de Matemática Aplicada III, Universidad Politécnica de Cataluña, Barcelona
marta.ginovart@upc.edu

²Dept. de Ingeniería Agroalimentaria y Biotecnología, Universidad Politécnica de Cataluña, Barcelona

Resumen

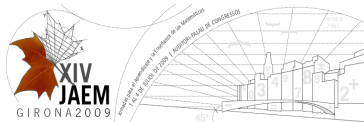
La actividad objeto de estudio se enmarca en una asignatura de estadística de nivel universitario. En este contexto, nuestro interés se centra en desarrollar un recurso para la actividad diaria de clase, que sea próximo a nuestro entorno y que este basado en datos reales, sobre objetos e instrumentos conocidos, situaciones de la actividad cotidiana, y sucesos recientes y/o frecuentes. Esta contribución trata sobre la generación y utilización de los resultados de una encuesta que los propios alumnos han ido respondiendo a lo largo de varios cursos académicos. La base de datos disponible actualmente ha facilitado el diseño de este recurso docente que permite integrar de forma natural tareas correspondientes a distintos temas estadísticos: descriptiva, probabilidad, variables aleatorias, métodos de inferencia y técnicas multivariantes.

PALABRAS CLAVE: estadística, encuesta, estudios universitarios

1. Introducción

Actualmente el desarrollo tecnológico ha puesto a nuestra disposición múltiples recursos que pueden ser utilizados para la enseñanza, el aprendizaje y la actividad estadística en el sentido amplio del término. Desde el análisis de los datos o estadística descriptiva, pasando por el estudio de las variables aleatorias y los modelos de probabilidad, para llegar a la utilización de métodos de estadística inferencial y métodos multivariantes, son muchas las herramientas informáticas que se han desarrollado y que ofrecen un apoyo indiscutible a la impartición de esta materia. No obstante, creemos que es necesario encontrar aplicaciones interesantes y adecuadas para utilizar y manipular en la actividad diaria de clase. Estas aplicaciones deberían proporcionar al estudiante la motivación necesaria para continuar con el estudio de la estadística durante todo el curso, y al mismo tiempo permitirle poner en práctica los conocimientos y las competencias que va adquiriendo. Por otra parte, en una primera etapa del proceso de aprendizaje debería de ser fundamental centrarse en aplicaciones próximas a nuestro entorno, basadas en datos reales, sobre objetos e instrumentos conocidos, situaciones de la actividad cotidiana, y sucesos recientes y/o frecuentes, con el fin de facilitar una aproximación realista al rol que tiene la estadística y a sus diversas posibilidades de uso. Esta contribución trata sobre la presentación de un recurso didáctico que responde a estas características y que se puede enmarcar en asignaturas de estadística correspondientes a primeros cursos universitarios.

La experiencia docente acumulada en estos años ha sido llevada a cabo en la Escuela Superior de Agricultura de Barcelona (ESAB) de la Universidad Politécnica de



Cataluña (UPC). En estos momentos la ESAB da formación en primero, segundo y tercer ciclo, en el ámbito de la ingeniería de los sistemas vivos. La asignatura de *Estadística* en la que se ha aplicado la propuesta desarrollada, es una asignatura troncal y obligatoria que se imparte en las tres titulaciones de ingeniería técnica agrícola de la ESAB (Explotaciones Agropecuarias, Hortofruticultura y Jardinería, Industrias Agrarias y Alimentarias) durante el segundo año (en tercer o cuarto semestre). Tanto para la *Estadística* como para el resto de las asignaturas del área de matemática aplicada de estas titulaciones, *Matemáticas I* y *Matemáticas II*, han sido diversas las actuaciones docentes implementadas durante los últimos años que sintonizan con el espíritu de mejora e innovación en la docencia, y con las directrices del proceso de Bolonia o Espacio Europeo de Educación Superior (Blanco y Ginovart 2007, 2008, 2009). Entendemos que ésta es una manera de ir adquiriendo experiencia en la adaptación de los estudios universitarios a un nuevo contexto condicionado por las necesidades y características actuales de nuestros estudiantes, así como para poder garantizar la consecución de las competencias genéricas y específicas que las nuevas titulaciones universitarias exigirán para un futuro próximo (Ginovart et al. 2008).

Es indispensable por tanto una adaptación progresiva en el ámbito universitario hacia la promoción de un modelo educativo que focalice la actividad docente en la mejora del aprendizaje centrado en el estudiante. Uno de nuestros fines en la práctica diaria docente dentro de la asignatura *Estadística* de la ESAB es conseguir un proyecto académico, que permita trabajar con éxito todas aquellas competencias genéricas y específicas que se le pueden atribuir a una asignatura de introducción a la estadística o estadística básica en los primeros cursos universitarios (Daniel, 2002). Este proyecto se ha ido gestando y configurando entorno a las respuestas que los propios alumnos han venido dando a una encuesta anónima durante los últimos ocho cursos académicos. Los estudiantes, durante el desarrollo cuatrimestral de la asignatura, elaboran un trabajo-proyecto que se estructura en distintas partes y que trata sobre el análisis de los resultados de esta encuesta, un conjunto de datos reales que ellos mismos han generado. La base de datos disponible actualmente ha facilitado la creación de un “recurso docente” que proporciona una manera de integrar de forma natural actividades o tareas correspondientes a los distintos temas de la asignatura, desde descriptiva, probabilidad, variables aleatorias, métodos básicos de inferencia estadística, hasta llegar a la introducción al análisis multivariante. A la vez, este recurso es utilizado como un hilo conductor en el desarrollo de la asignatura y facilita su ampliación. El objetivo de esta contribución es presentar los elementos utilizados en la preparación de este recurso didáctico, su presencia en las distintas partes de la asignatura con el fin de ilustrar aquellos procesos cognitivos que pretende estimular y practicar, los cambios metodológicos en el aula vinculados a su uso, así como los resultados que ha proporcionado en los diferentes cursos académicos donde progresivamente se ha ido implementando.

2. Metodología

2.1. Contexto de la asignatura de *Estadística*

Con el objetivo de clarificar y afianzar la prioridad de los contenidos de la asignatura de *Estadística*, en el curso 2007/2008 se realizó una encuesta a 50 profesores implicados

en la docencia de las actuales titulaciones y másters oficiales de la ESAB. En ese sondeo se pidió cuáles, de entre una lista de ítems estadísticos, eran los temas que consideraban que debían formar parte de la instrucción “estadística” de un estudiante. Las respuestas obtenidas son las procuradas por 50 profesores de asignaturas pertenecientes a materias muy diversas: Ciencias del medio natural, Ciencias y Tecnología del medio ambiente, Economía, Ingeniería, Operaciones Básicas en Industrias Agroalimentarias, Producción animal, Proyectos, Química, Tecnología de la producción vegetal y Física entre otras. Los resultados de este sondeo se muestran en la Figura 1. De las respuestas recogidas cabe destacar que un porcentaje alto de los encuestados, alrededor de 70%, considera muy importantes las cuestiones relacionadas con la obtención de datos experimentales y su análisis descriptivo, métodos básicos de inferencia estadística, así como el manejo de programas estadísticos. Un porcentaje menor, alrededor del 30%, se inclina por la inclusión en este contexto de métodos de estadística multivariante. Es indiscutible que estos estudios de ingeniería requieren de un bagaje estadístico significativo, y que debemos poner especial énfasis en conseguir que el estudiante sea capaz de combinar una visión intuitiva de los problemas a resolver con la precisión del lenguaje propio del ámbito estadístico y la utilización de los métodos y/o procedimientos adecuados.

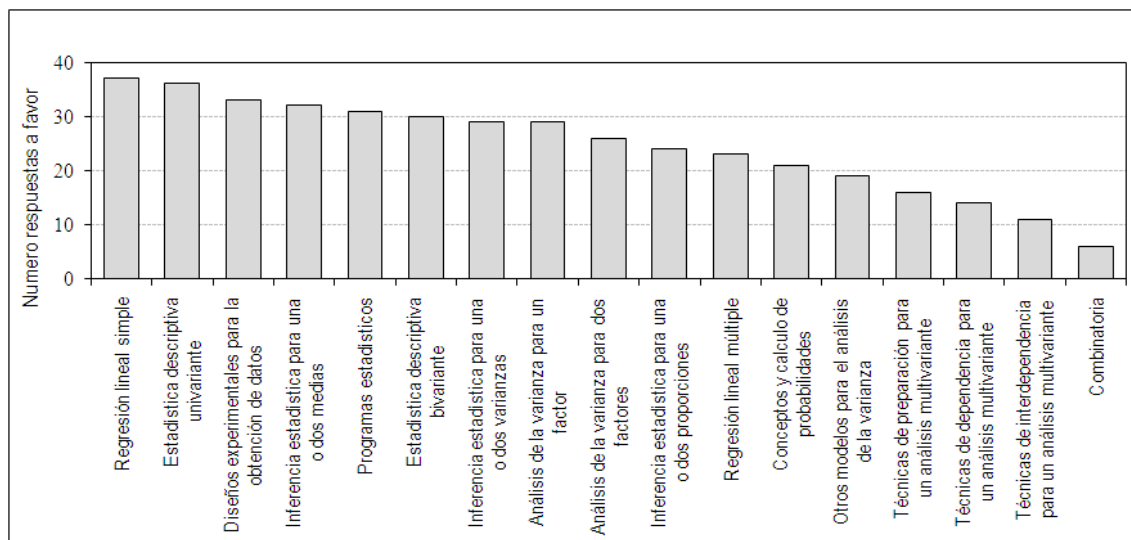


Figura 1. Resultados de la encuesta realizada en el curso académico 2007/2008 a 50 profesores que imparten docencia en los estudios de ingeniería técnica agrícola y másters de la ESAB.

2.2. Diseño de la encuesta y respuestas para analizar

En el curso 2000/2001 surgió la idea de preparar una encuesta para pasarla a los estudiantes al inicio de la asignatura de *Estadística* con la intención de que los resultados que se obtuvieran fuesen analizados por ellos mismos. Desde entonces, y de forma regular, se han ido recogiendo las respuestas anónimas de esa encuesta, la cual no ha sufrido cambios en su formato y contenido, con el fin de ir ampliando la base de datos sobre la cual trabajar. La Tabla 1 presenta las 21 preguntas de la encuesta. En el diseño de la encuesta se optó por un conjunto de preguntas diversas sobre la persona del estudiante (preguntas 1, 2, 15, 16), sus gustos o preferencias (3, 4, 5, 8, 21), su rendimiento académico (10, 11, 12) así como información relativa a sus actividades

fuera del contexto académico (7, 17, 18) o relativas a su entorno socioeconómico y cultural (6, 9, 13, 14, 19, 20). El conjunto de respuestas a obtener consiste en una colección de datos propios y variados, que responden a una realidad bien conocida por todos los estudiantes, que les permite investigar diversos aspectos en función de su interés personal e inquietudes.

Tabla 1. Formato de la encuesta que se pasa a los estudiantes al iniciar la asignatura de *Estadística*.

Encuesta 2009 Utilizarás los resultados de esta encuesta para realizar el proyecto de la asignatura.		
1. Sexo	1 Hombre, 2 Mujer	1 - 2
2. Edad actual (en años)		
3. ¿Cuántos cigarrillos fumas diariamente?	(0 si no fumas)	
4. ¿Cuánto tiempo dedicaste a navegar por Internet la semana pasada?	(en horas)	
5. ¿Cuánto tiempo dedicaste a ver la TV la semana pasada?	(en horas)	
6. ¿Cuál es el gasto mensual de tu móvil?	(0 si no tienes móvil o no lo utilizas)	
7. ¿Cuánto tiempo dedicaste la semana pasada a hacer deporte?	(en horas)	
8. ¿Cuál de estos equipos de fútbol prefieres?	1-Barcelona 2-Madrid 3-Español 4-Ninguno de los anteriores	1 - 2 - 3 - 4
9. Procedencia:	1-COU 2-LOGSE 3-FP 4-Mayores de 25 años 5-Otra	1 - 2 - 3 - 4 - 5
10- Nota de acceso a estos estudios (en número entero)		
11- Nota de la asignatura <i>Matemáticas I</i>	1-Suspenso 2-Compensable 3-Aprobado 4-Notable 5-Excelente 6-MH	1 - 2 - 3 - 4 - 5 - 6
12- Nota de la asignatura <i>Matemáticas II</i>	1-Suspenso 2-Compensable 3-Aprobado 4-Notable 5-Excelente 6-MH	1 - 2 - 3 - 4 - 5 - 6
13- ¿Qué tipo de transporte utilizas para venir a la Universidad?	1-Transporte público 2-Coche 3-Moto 4-A pie 5-Bicicleta 6-Otro (si utilizas más de uno, indica aquel con el que el trayecto es el más largo o es el más habitual)	1 - 2 - 3 - 4 - 5 - 6
14- ¿Cuánto tiempo tardas normalmente en llegar a la Universidad desde tu casa?	(en minutos)	
15- Peso (en kg)		
16- Altura (en cm)		
17- Situación laboral actual	1-Trabajas 2-No trabajas	1 - 2
18- Importe del sueldo y/o ingresos mensuales totales que no provengan de asignaciones familiares (en euros)		
19- Nivel de formación del padre	1-Sin estudios 2-Estudios primarios 3-Estudios secundarios 4-Estudios superiores	1 - 2 - 3 - 4
20- Nivel de formación de la madre	1-Sin estudios 2-Estudios primarios 3-Estudios secundarios 4-Estudios superiores	1 - 2 - 3 - 4
21- ¿Qué deporte prefieres practicar?	1-Fútbol 2- Baloncesto 3-Natación 4- Tenis 5- Windsurf 6-Otro	1 - 2 - 3 - 4 - 5 - 6

Desde el curso 2000/2001 se han ido organizando y almacenando las respuestas de los estudiantes de la asignatura *Estadística* correspondiente a cada una de las tres titulaciones de ingeniería técnica agrícola en las que hemos impartido docencia (Tabla

2). Actualmente la base de datos aglutina las respuestas de cerca de 600 encuestas. La información muestral corresponde a 23 variables distintas, las correspondientes a las 21 preguntas formuladas de la Tabla 1, que serán designadas como variables P1, P2, ..., y P21, junto con las dos variables que identifican el curso académico y la titulación (CA y T, respectivamente). Es decir, los alumnos del curso actual 2008/2009 disponen de aproximadamente unos 13000 datos para analizar. La Figura 2 muestra cómo se han organizado los resultados de las encuestas en una hoja de cálculo, que los estudiantes tienen a su disposición en la plataforma Atenea (entorno virtual de soporte a la docencia de la UPC), con el fin de que puedan manipular todo el conjunto de datos. El análisis de los datos puede hacerse directamente con la propia hoja de cálculo, o bien con los programas estadísticos que la asignatura presenta y que se encuentran operativos en las aulas informáticas del centro (Minitab y SAS).

Tabla 2. Número de estudiantes que han respondido la encuesta.

	Explotaciones Agropecuarias	Hortofruticultura y Jardinería	Industrias Agrarias y Alimentarias
Curso 2000/2001	33	38	-
Curso 2001/2002	40	28	32
Curso 2002/2003	28	42	-
Curso 2003/2004	-	35	13
Curso 2004/2005	-	46	34
Curso 2005/2006	26	27	-
Curso 2006/2007	29	32	-
Curso 2007/2008	21	39	-
Curso 2008/2009	17	25	-

</

Figura 2. Detalle de la base de datos elaborada con las encuestas respondidas por los estudiantes.

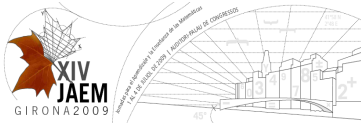
2.3. Actividades para analizar los resultados de la encuesta

A la hora de plantear los proyectos o trabajos en una asignatura con frecuencia surge la cuestión de cómo organizar el trabajo individual o en equipo de los estudiantes. La respuesta a la pregunta de si este trabajo debe de evolucionar de forma autónoma, o si es más conveniente que el profesor haga las indicaciones pertinentes y concretas para cubrir cada una de las etapas, dependerá del contexto y de los objetivos que se quieran alcanzar. Este análisis de datos puede ser guiado de forma específica en el inicio de los temas para ilustrar o practicar los conceptos y métodos expuestos en clase. Como un trabajo global y de síntesis de la asignatura, o de los distintos módulos de ésta, debe surgir de forma natural, como una consecuencia de los intereses o inquietudes personales de los estudiantes frente a la información que se les proporciona, y con la que se puedan encontrar directamente implicados. En los distintos cursos impartidos en los que hemos estado utilizando estos datos hemos ido ensayando las distintas opciones, tanto la de tareas localizadas y dirigidas como la de un trabajo abierto y de contenido libre. Con la experiencia acumulada en cursos sucesivos, se han ido preparando ejercicios para realizar de forma específica, y se han ido formulando cuestiones para orientar en la opción de trabajo más autónomo. El fin último era conseguir enmarcar el conjunto de todos estos análisis estadísticos en un proyecto académico completo, que permitiera tratar los distintos temas que configuran actualmente la asignatura *Estadística* en nuestro centro.

El conjunto de datos reúne tanto variables cualitativas, en escala nominal (CA, T, P1, P8, P9, P13, P17, P21) y ordinal (P11, P12, P19, P20), como variables cuantitativas discretas (P2, P3, P10) y continuas (P4, P5, P6, P7, P14, P15, P16, P18). Esto permite realizar análisis estadísticos distintos atendiendo a criterios diversos y utilizando estrategias múltiples. Si consideramos la relación de objetivos de una asignatura de estadística, una parte muy importante de éstos se puede lograr con una manipulación efectiva y adecuada de estos datos.

En primer lugar, en la parte de estadística descriptiva, para investigar o responder preguntas o cuestiones diversas relacionadas con las variables de la base de datos de la encuesta, se pueden aplicar los siguientes procedimientos:

- ✓ Clasificar variables, en cualitativas, cuantitativas discretas y cuantitativas continuas, y definir subgrupos de datos.
- ✓ Decidir si agrupar o no un conjunto de datos en intervalos y escoger el número de intervalos más adecuado.
- ✓ Construir tablas de frecuencias de conjunto de datos (agrupados o no).
- ✓ Para conjuntos de datos univariantes, identificar el tipo de gráfico más conveniente según la naturaleza de la variable, y representarlo gráficamente utilizando: diagrama de Pareto, diagrama de puntos, diagrama de barras, diagrama de sectores, diagrama de tallo y hojas, histograma, gráfico poligonal, o diagrama de caja.
- ✓ Calcular y discutir los valores de los estadísticos descriptivos de tendencia central (media aritmética, mediana, moda), de posición (cuartiles, percentiles), de dispersión (recorrido, rango intercuartílico, varianza, desviación típica, coeficiente de variación) y de forma (coeficiente de asimetría, curtosis)
- ✓ Identificar y discutir la presencia de datos atípicos para las diferentes variables.



- ✓ Representar de forma conjunta datos bivariantes cuantitativos e interpretar su relación.
- ✓ Organizar tablas de doble entrada o de contingencia, e interpretar los valores obtenidos con la ayuda de la representación gráfica de los diferentes perfiles conseguidos.

En segundo lugar, en la parte de probabilidad y variables aleatorias podemos usar los datos de la encuesta para realizar algunas de las aplicaciones prácticas siguientes:

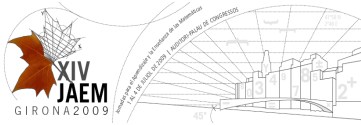
- ✓ Valorar la dependencia e independencia de ciertos sucesos o a partir de las tablas de contingencias creadas anteriormente, o con los gráficos de determinados perfiles.
- ✓ Valorar si una distribución de un conjunto de datos se ajusta a una distribución normal o aproximadamente normal (aplicando la regla 68-95-99.7 o bien con gráficos de probabilidad).
- ✓ Discutir y valorar si una determinada variable sigue alguna de las distribuciones de probabilidad siguientes: uniforme discreta, binomial, Poisson, uniforme continua o exponencial.

Finalmente en la parte de inferencia estadística básica, en relación con los procedimientos de estimación y contraste de hipótesis el conjunto de datos nos permite esbozar el papel que tiene todo el conjunto completo de datos, asimilable al de una “población”, y el rol que pueden tener subconjuntos de datos como muestras aleatorias simples (pequeñas o grandes) de la población, para poder:

- ✓ Utilizar la estimación puntual y la estimación por intervalos de confianza de parámetros poblacionales como la esperanza, la varianza y la proporción.
- ✓ Contrastar pruebas de significación para parámetros poblacionales como la esperanza, la varianza y la proporción.
- ✓ Comparar dos parámetros poblacionales, a partir de medias muestrales, varianzas muestrales y proporciones muestrales, utilizando intervalos de confianza o pruebas de significación.
- ✓ Aplicar el test de independencia Chi-cuadrado para dos variables con las tablas de contingencia obtenidas.

En relación a la aplicación de los modelos lineales para algunas parejas de variables se puede obtener la recta de regresión de mínimos cuadrados, analizar la validez del modelo de regresión lineal simple, y llevar a cabo la inferencia estadística a partir de este modelo, así como ensayar algún modelo de regresión múltiple en casos de extensión de temario.

En el contexto de la asignatura optativa *Diseño experimental y obtención, validación y tratamiento de datos* del máster “Sistemas Agrícolas Periurbanos” que se imparte en la ESAB, se utilizaron los resultados de esta encuesta como ejercicio de aplicación para mostrar una técnica multivariante como es la del análisis factorial (Hair et al. 1999). El propósito final era definir la estructura subyacente de la matriz de datos, para poder conseguir un resumen y una reducción de ésta, encontrando la manera de condensar la información contenida en la serie de variables originales en una serie más pequeña, de dimensiones compuestas o valores teóricos nuevos (factores), con una mínima pérdida de información. Siendo una técnica de interdependencia en la que se consideran todas las variables simultáneamente, cada una relacionada con todas las demás, y donde los factores son combinaciones lineales de las variables, es muy apropiada para los resultados de la encuesta recogidos. Es una técnica interesante de aplicar a estos datos,



ya que los factores se forman para maximizar su explicación en toda la serie de variables, extrayendo lo que proporcionan los datos, y a priori, sin hacer ninguna restricción inicial sobre la estimación de los factores o sobre el número de factores a considerar. Desde el punto de vista del estudiante, este tipo de análisis constituye una buena oportunidad para tener una visión conjunta de la información recogida. A la vez, la revisión de los requerimientos que hace que esta técnica sea adecuada, implica hacer acciones manipulativas diversas sobre la base de datos. Por una parte, es necesaria la adecuación o reformulación de variables de escala ordinal a escala de razones o proporciones para que puedan ser incorporadas en el estudio, y por otra, la realización del análisis de los valores perdidos o preguntas no contestadas por algunos estudiantes.

3. Resultados y discusión

En relación a la opción de utilizar estos datos de forma dirigida por el profesor, nuestra experiencia en el aula nos indica que es buena en una primera fase del proceso de aprendizaje. Sin embargo, nos inclinamos por la opción de un trabajo más abierto, sin excesivas pautas para su ejecución, ya que esta opción requiere de una mayor autonomía por parte del estudiante y, por tanto, le exigirá para que tenga éxito una mayor madurez “estadística”. Esta segunda opción demanda una visión más global de las posibilidades gráficas, de los resúmenes numéricos y de los tests estadísticos, ya que exige que el estudiante decida, que tenga que escoger y priorizar unas manipulaciones y ejecuciones específicas frente a otras, pudiendo mostrar su determinación, y reflejar su iniciativa y creatividad en este contexto. Esto implica realizar un razonamiento crítico y riguroso de la situación y de los objetivos que quiera alcanzar o de las hipótesis que quiera contrastar, así como de los resultados que finalmente considere interesantes de mostrar. Esta segunda opción se acerca mucho más a lo que será su práctica estadística futura, en trabajos o informes para otras asignaturas o bien en el análisis de los datos experimentales de su trabajo final de carrera.

Hemos observado que hacer referencias frecuentes a las variables de la encuesta cuando se presentan y desarrollan los nuevos conceptos o procedimientos estadísticos en clase facilita que los alumnos se familiaricen con todo el conjunto de datos, recojan las ideas posibles para los futuros análisis de estos datos, y aumenten las propuestas creativas y originales en la ejecución práctica de estos análisis.

La mayor utilización de los datos de la encuesta se ha desarrollado en el módulo de estadística descriptiva, aunque actualmente se está extendiendo progresivamente a los otros módulos o partes de la asignatura. En las Figuras 3-10 mostramos algunos resultados presentados y discutidos por los estudiantes en los últimos cursos académicos. Éstos han sido obtenidos con indicaciones específicas proporcionadas por el profesor, o bien escogidos de algunos trabajos entregados por los estudiantes en respuesta a la pregunta “¿Qué aspectos te parecen más interesantes de analizar de la base de datos confeccionada con las respuestas de la encuesta realizada?”. En ese último caso, de forma individual o en grupo de 2 o 3 estudiantes, ellos decidían qué variables (hasta un máximo de 5) iban a investigar, cómo lo iban a hacer, y con qué finalidad.

En el curso 2006/2007 el proyecto se llevó a cabo en grupos formados por tres estudiantes. Con el fin de supervisar el trabajo realizado por cada miembro del equipo,

se realizó una evaluación individual del proyecto correspondiente a la parte descriptiva de la asignatura que consistía en responder las siguientes preguntas y cuestiones: “¿Por qué seleccionasteis las variables con las que habéis trabajado?”, “¿Cómo planteasteis el análisis de los datos?”, “¿Con qué dificultades os encontrasteis?”, “Describe de forma concisa algunas de las herramientas metodológicas utilizadas en vuestro proyecto”, y “Comenta brevemente algunos de los resultados obtenidos que te hayan parecido interesantes o que te hayan llamado la atención”. Realmente, las respuestas que dieron la mayoría de los estudiantes fueron muy correctas, originales, elaboradas y sorprendentemente bien argumentadas, con sentido y criterio, reflejando que habían entendido bien las posibilidades de un buen análisis de datos reales.

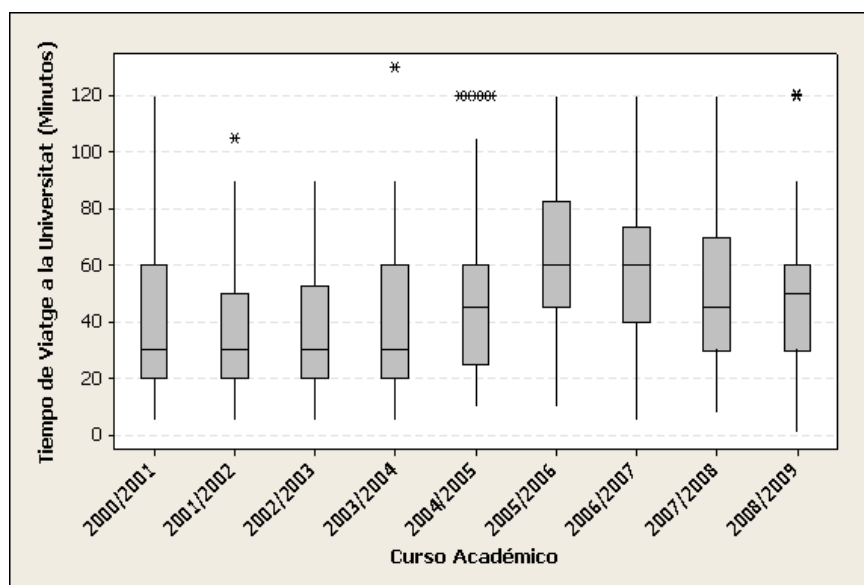
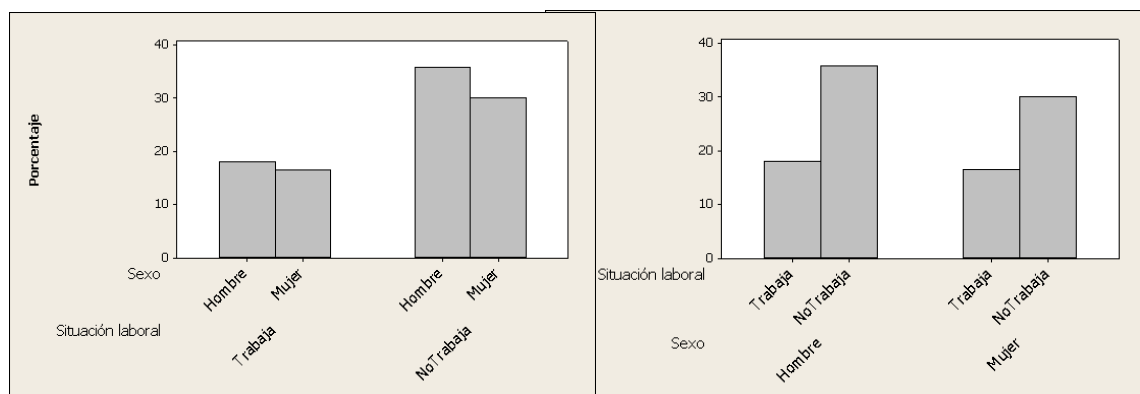


Figura 3. Diagrama de caja proporcionado por el Minitab para la variable P14 “Tiempo invertido en el trayecto para ir a la universidad” por los estudiantes según la variable CA “Curso académico”.



Salida del Minitab: Pearson Chi-Square = 0,198; DF = 1; P-Value = 0,657

Figura 4. Estudio de la independencia entre la variable P1 “Sexo” y la P17 “Situación laboral” de los estudiantes.

Formación padre					Formación madre				
	Count	CumCnt	Percent	CumPct		Count	CumCnt	Percent	CumPct
1	16	16	2,77	2,77	1	21	21	3,63	3,63
2	162	178	28,03	30,80	2	204	225	35,29	38,93
3	163	341	28,20	59,00	3	145	370	25,09	64,01
4	237	578	41,00	100,00	4	208	578	35,99	100,00
N=	578				N=	578			

Figura 5. Salida del Minitab de tablas de distribuciones de frecuencias para las variables P19 y P20 “Nivel de formación del padre” y “Nivel de formación de la madre”, donde 1: Sin estudios, 2: Estudios primarios, 3: Estudios secundarios, 4: Estudios superiores.

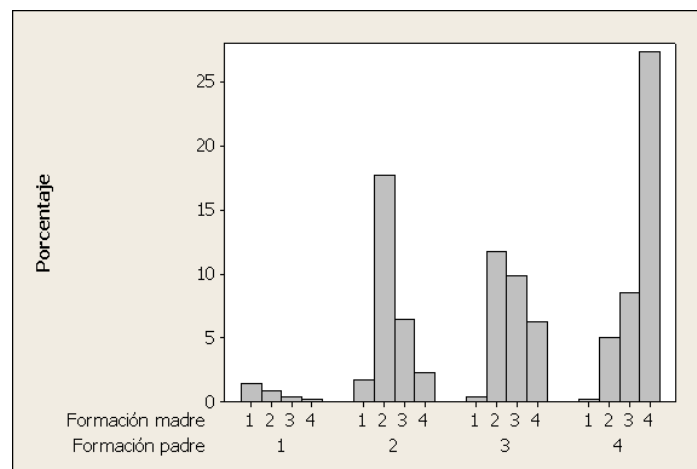


Figura 6. Diagramas de barras para los perfiles de la distribución de la variable P20 “Nivel de formación de la madre” según los distintos niveles de la variable P19 “Nivel de formación del padre”. 1-Sin estudios 2-Estudios primarios 3-Estudios secundarios 4-Estudios superiores.

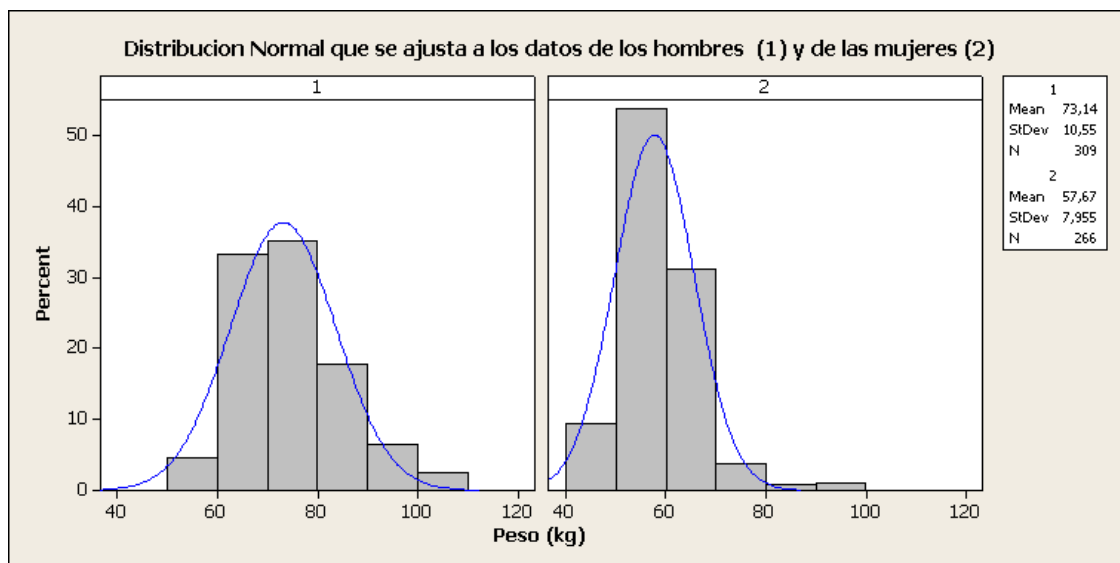


Figura 7. Salida del Minitab para valorar la aproximación de la distribución de la variable P15 “Peso” de los estudiantes clasificados por la variable P1 “Sexo”.

Rows: Matemáticas I		Columns: Matemáticas II		
S-C:Suspense o Compensable A:Aprobado N-E-MH:Notable o Excelente o MH				
	S-C	A	N-E-MH	All
S-C	60	60	15	135
	44,44	44,44	11,11	100,00
	42,86	19,48	11,45	23,32
	10,36	10,36	2,59	23,32
A	72	221	62	355
	20,28	62,25	17,46	100,00
	51,43	71,75	47,33	61,31
	12,44	38,17	10,71	61,31
N-E-MH	8	27	54	89
	8,99	30,34	60,67	100,00
	5,71	8,77	41,22	15,37
	1,38	4,66	9,33	15,37
All	140	308	131	579
	24,18	53,20	22,63	100,00
	100,00	100,00	100,00	100,00
	24,18	53,20	22,63	100,00
Cell Contents:	Count			
	% of Row			
	% of Column			
	% of Total			
Pearson Chi-Square = 118,851; DF = 4; P-Value = 0,000				

Figura 8. Salida del Minitab para el estudio de la independencia entre la variable P11 “Nota Matemáticas I” y la P12 “Nota Matemáticas II” de los estudiantes.

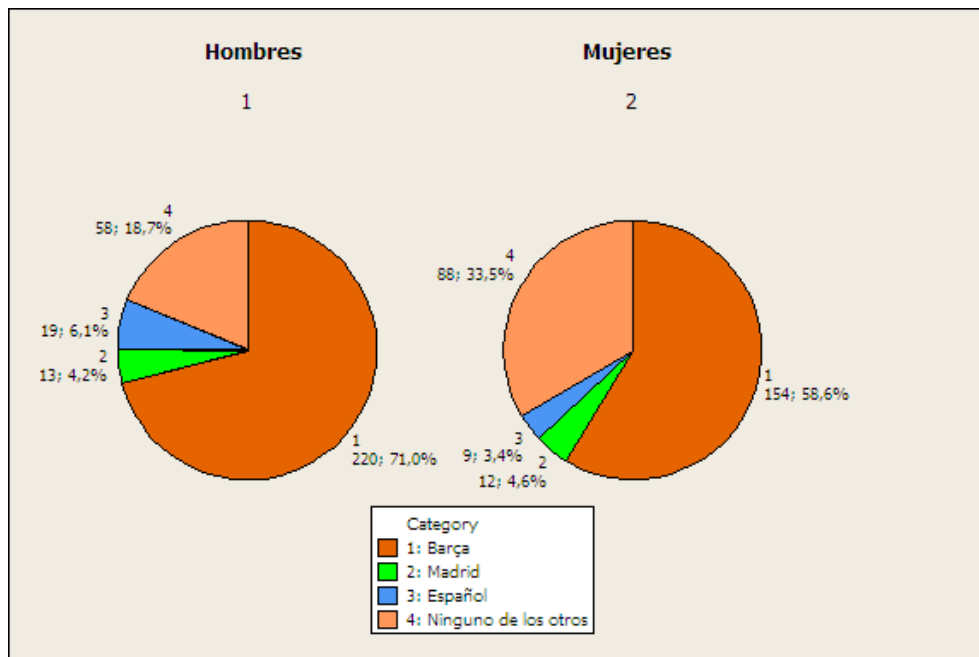


Figura 9. Gráfico de sectores de la variable P8 “Preferencia equipo de fútbol” en función de la variable P1 “Sexo”.

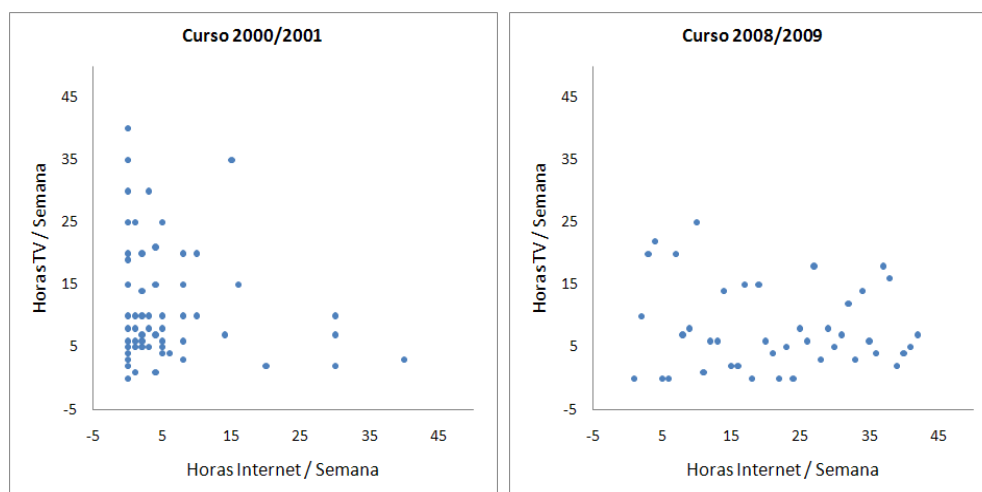
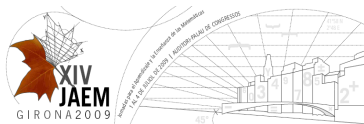


Figura 10. Gráficos de dispersión correspondientes a las variables P4 “Tiempo dedicado a navegar por Internet” y P5 “Tiempo dedicado a ver la TV” en dos cursos académicos.

En el curso 2007/2008, al terminar el módulo de probabilidad y variables aleatorias de la asignatura *Estadística* se propuso una actividad pautada para abordar la aplicación de algunos de los conceptos trabajados en el aula a la base de datos obtenida. Se entregó a los estudiantes una hoja de papel con un análisis estadístico descriptivo completo de los datos correspondientes a algunas de las preguntas de la encuesta, y se les preguntaba sobre la posibilidad de que las distribuciones obtenidas siguieran determinados modelos de probabilidad (Bernoulli, binomial, Poisson, geométrica, uniforme, exponencial o normal). Dada la disponibilidad de información y recursos con que contaban en el momento de realizar la actividad, resúmenes gráficos y numéricos, pero no de un programa estadístico, las respuestas no podían ser cerradas y únicas, sino que ofrecían la oportunidad de discutir y sopesar los argumentos a favor y los argumentos en contra, respecto a la distribución empírica de la variable y a la posible adecuación a una distribución teórica de probabilidad ya estudiada. Por tanto era necesario realizar cálculos aproximados para discutir la conveniencia de un posible modelo probabilístico. La Figura 11 muestra un detalle del material entregado al estudiante para esta actividad. Al final de ésta se propuso a los estudiantes que la valorasen (de forma anónima), indicando cuál era su opinión. La Figura 12 muestra los resultados de la valoración de esta actividad de los 59 estudiantes que la realizaron.

Considerando la posibilidad de extracción de muestras aleatorias simples de todo el conjunto de datos o bien de subgrupos determinados, algunos estudiantes han incorporado en sus trabajos al finalizar la asignatura la utilización de los métodos de estimaciones por intervalos, los test z y t de Student para contrastar hipótesis relativas a medias y proporciones, y el test F de Fisher para varianzas. Las pruebas de significación que se realizaron fueron básicamente guiadas por los resultados observados en el estudio descriptivo previo de los datos. No obstante es una de las partes de la actividad que debería ser potenciada, ya que no es tan fácilmente abordable para los estudiantes como lo es la parte de descriptiva. Para potenciar más la utilización de datos desde la perspectiva inferencial, como trabajo futuro nos proponemos diseñar e implementar algunas tareas específicas con esta base de datos.

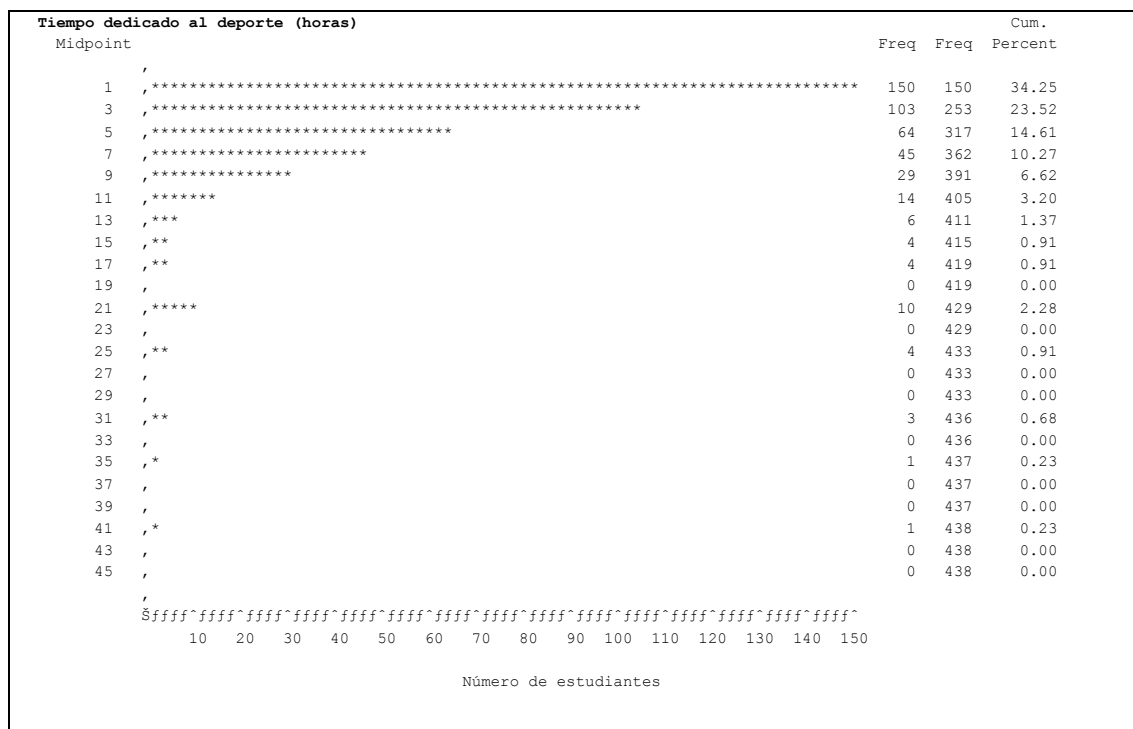
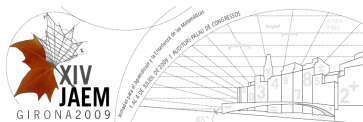


Figura 11. Salida del SAS que representa un histograma para la variable P7 “Tiempo semanal de deporte”.

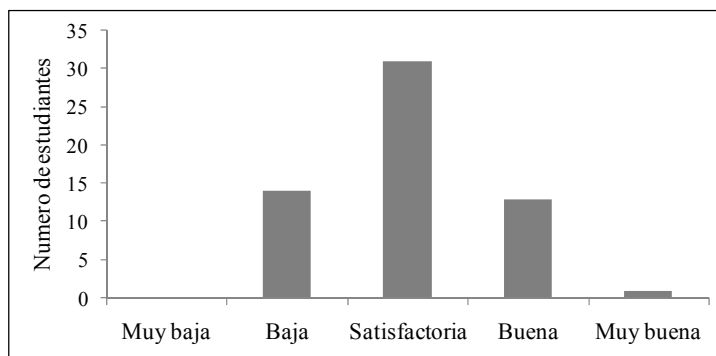


Figura 12. Valoración de los 59 estudiantes del curso 2007/2008 de la actividad realizada con datos de la encuesta y distribuciones de probabilidad.

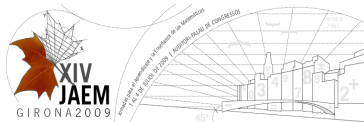
En la ampliación de estadística y en relación a la aplicación de la técnica de análisis factorial (tipo R) a este conjunto de datos, la manipulación previa para poder transformar las escalas utilizadas en las variables y el estudio de los valores perdidos, obligó al estudiante a una búsqueda adicional de información y de procedimientos del programa estadístico SAS. Estos requerimientos previos y necesarios para un análisis factorial de datos, permitió familiarizarse con sentencias de lógica condicional (*if ... then ... else*) y programación (Figura 13). El estudio de los valores perdidos o de no respuesta se tuvo que realizar a dos niveles, primero identificando su número por variable y después estudiando posibles patrones específicos en su disposición. Esta tarea requirió de técnicas descriptivas diversas. Al tratarse de un problema no “preparado” expresamente para ello, los resultados que se consiguieron no fueron muy concluyentes.

Esto, no obstante, fue valorado positivamente por el estudiante que realizó el trabajo, ya que era necesario ser crítico y prudente en el momento de hacer la valoración de los resultados que relacionaban aspectos bien conocidos por él. Evidenció también la necesidad de ser consciente de las limitaciones que todo método estadístico tiene, y que los datos en este caso reflejaban. Se tuvo la oportunidad de comprobar que si en un análisis de datos se incluyen indiscriminadamente grandes cantidades de variables y uno espera que la estadística “lo solucione”, entonces la posibilidad de obtener “malos resultados” es alta. A pesar de que el estudio no se extendió a otras técnicas multivariantes quedó patente la posibilidad de realizar otros estudios identificando variables discriminantes. Ésta es otra de las tareas pendientes que esperamos poder abordar en un futuro próximo en el contexto de ampliación de estadística.

```
data kk;
infile 'resulencuesta20072008.txt' delimiter='09'x firstobs=2;
input curs$ especialitat$ sexe edat ..... fmare esportp;
options nodate;
*CREACION DE VARIABLES PARA EL ESTUDIO DE LOS RESIDUS;
if sexe= . then nsexe = 0; else nsexe=1;
if edat= . then nedat = 0; else nedat=1;
....
if nsexe=0 or nedat=0 or ..... or nesportp=0 then perdutstotals=0;
else perdutstotals= 1;
*TRANSFORMACION DE NOTAS DE ESCALA DESCRIPTIVA A ESCALA NUMERICA;
if mates1= 1 then mates11= 2.5;
if mates1= 2 then mates11= 4.5;
....
proc means data= kk n nmiss sum;
ods html file="estudiperduts.xls.";
proc print data = kk;
var curs especialitat nsexe nedat ..... nfmare nesportp;
ods html close;
*EXAMEN CORRELACIONES MAYORES 0,3 Y SIGNIFICACION;
proc corr data=kk;
var edat cigarrets ..... ttransport pes altura ingressos;
*EXAMEN CORRELACIONES PARCIALES Y MSAs;
proc factor data=kk corr method=principal msa;
var edat cigarrets tinternet ..... ttransport pes altura ingressos;
run;
*SACAR TTRANSPORT;
proc factor data=kk corr method=principal msa;
var edat cigarrets tinternet ..... pes altura ingressos;
....
*ANALISI CON 7 VARIABLES;
proc factor data=kk corr method=principal msa;
var tinternet ttv tesport notaacces mates11 mates22 altura;
proc factor data=kk out=solucio reorder corr method=principal msa preplot plot
rotate=varimax scree eigenvectors nfactors=2;
var tinternet ttv tesport notaacces mates11 mates22 altura;
proc print data = solucio;
proc gplot data=solucio;
plot factor1*factor2 = sexe /Vref=0 href=0 haxis = -3 to 3 by 1 vaxis = -3 to 3 by 1;
symbol value = dot;
proc gplot data=solucio;
plot factor1*factor2 = especialitat /Vref=0 href=0 haxis = -3 to 3 by 1 vaxis = -3 to 3
by 1;
symbol value = dot;
run;
```

Figura 13. Detalle del programa que se ejecutó en SAS para llevar a cabo el análisis factorial de los datos.

En resumen, la manipulación de estos datos ha obligado a los estudiantes a utilizar distintos soportes informáticos que permiten agilizar cálculos y aplicar diversos métodos de inferencia estadística. Como consecuencia de ello, las salidas proporcionadas por el ordenador han requerido de una correcta interpretación y consistencia para llegar a presentar y discutir los resultados de forma global y conjunta.



Con ayuda de la hoja de cálculo Excel y de los programas estadísticos Minitab y SAS, los estudiantes han sido capaces de organizar un gran conjunto de datos originales, llevar a cabo análisis descriptivos y procedimientos estadísticos de estimación puntual o test de hipótesis esbozados por ellos mismos y que respondían a sus intereses particulares. La generación y utilización de esta encuesta ha resultado ser una recurso docente muy apto para la asignatura *Estadística*, pues los estudiantes han constatado que es necesario conocer bien tanto los datos como las variables que se relacionan para interpretar mejor los resultados. Se han sentido muy involucrados en todo este proceso de análisis ya que han tratado y analizado sus propios datos, lo cual ha permitido combinar sus inquietudes personales con la manera de “hacer estadística”, manteniendo o, incluso, incrementando, su interés por la asignatura.

4. Agradecimientos

Queremos agradecer a los estudiantes que participaron en la generación de estos datos.

5. Bibliografía

- Blanco, M., Ginovart, M. (2007). Suggestions for Teaching Exploratory Data Analysis in Engineering Education within the European Credit Transfer System. En: J. Szendrei (Eds), *Mathematical activity in classroom practice and as research object in didactics: two complementary perspectives*, pp. 65-72. Budapest: Varga Tamás Foundation.
- Blanco, M., Ginovart, M. (2008). La probabilidad en las enseñanzas técnicas dentro del marco del Espacio Europeo de Educación Superior, con apoyo de la plataforma Moodle. En R.B. Bárcena, J.J. Domínguez, A.Morgado, A. Quirós (Eds), *Actas del XVI Congreso Universitario de Innovación Educativa en las Enseñanzas Técnicas 16 CUIEET*. Cadiz: Universidad de Cádiz.
- Blanco, M., Ginovart, M. (2009). Introducing the normal distribution by following a teaching approach inspired by history: an example for classroom implementation in engineering education. *Actas del Sixth Conference of European Research in Mathematics Education CERME6*. En prensa.
- Daniel, W. W. (2002). Bioestadística. Base para el análisis de las ciencias de la salud. Limusa Wiley.
- Ginovart M., Blanco M., Portell X. (2008). Una reflexión sobre la actividad docente en el área de matemática aplicada en estudios universitarios de ingeniería de biosistemas: presente y futuro. *Actas del XII Congreso THALES de Enseñanza y Aprendizaje de las Matemáticas XII CEAM*. En prensa.
- Hair, J. F. Jr, Anderson, R. E., Tatham, R. L., Black, W. C. (1999). Análisis multivariante. Prentice Hall Iberia.